RESOURCE OPTIMIZATION IN CLOUD COMPUTING ENVIRONMENT USING SOFT COMPUTING TECHNIQUES

A Synopsis submitted in the partial fulfilment of the requirement for the award of the degree of

Doctor of Philosophy
in
Computer Science

Submitted by
Mohit Agarwal

Under the Supervision of
Dr. Gur Mauj Saran Srivastava

Prof. G.S. Tyagi
Head,
Dept. of Physics & Computer Sc.
Faculty of Science

Prof. R. Kumar
Dean,
Faculty of Science

DEPARTMENT OF PHYSICS AND COMPUTER SCIENCE
FACULTY OF SCIENCE
DAYALBAGH EDUCATIONAL INSTITUTE
DAYALBAGH
AGRA (UP) – 282005

1. Introduction

In this digitization era, the need for the accessing the resources over the internet like utilities was required very much. Now every organization is looking over the internet for their requirement to meet up to store and process the data.

In present scenario, Cloud computing presents itself as an efficient mechanism for the storage of data and data processing in third party sites. This involves the outsourcing and sharing of resources to acquire a greater efficiency in terms of economy, scalability and resource utilization. As now days, human being is getting more and more dependent on technology, there is a need for the utility which can be access on the move, as on demand and where and whenever it is required. The solution comes in the form of the cloud computing which allow the accessing of resources over the internet on the demand basis. Like human being can access the their daily required needs like water, power etc. now a day's more and more companies like Google's compute engine, Amazon's Web Services, IBM's softlayer and many are coming forward to offer the cloud service to the intended user. The idea is also getting the popularity as setting the own set up for the companies is proving as the more expensive investment in terms of hardware, software and the people resource to manage them so they are more and more interested in availing the same services what they required from the cloud services provider companies. Now the only primary resource they need is internet connectivity to access the services which may present anywhere across the globe. Now the startups and middle level organization no matter having Information Technology related knowledge or not can more focuses on their core business and devote much time on the core business related innovations using the available resources on payment basis by such services provider. Cloud computing has gained a remarkable popularity in industry and academia also [46]. The term "Cloud" may be explained as a cluster or pool of I.T. related resources like RAM, HD's, CPU etc which can be make available for the use of client as and when needed basis on the paid basis. But as the property of cloud the resources which are available for the users may not be fixed all the time and moreover the multiple users may be accessing the same resources at the time. So proper techniques are required to exploit the

resources fully and in efficient manner in terms of power consumption, processors utilization etc so that the existing set up will result more profitable.

In 1960's J.C.R. Licklider, turns with the idea of "intergalactic computer network" this led to the development of ARPANET (Advanced Research Projects Agency Network) in 1969. It's his vision to connect everybody and data from any site and anywhere on the globe using the network.

In 1960's John McCarthy was other scientist, who delivers the idea of utility computing into the technology world, serving the people or user with the data as per their demand like water and electricity [1].

In 1972 IBM developed the concept which they termed as VM Operating System which provides the time sharing services via VM's Virtual Machine.

Year 1999, brings the first key milestone in the practical area of cloud computing history can be arrival of salesforce.com who were masters in the delivery of enterprise solution via a simple website.

The main revolt in the field of cloud computing come with the introduction of amazons Elastic Cloud Compute which is popularly known as EC2 which is followed by many more organizations in the field of cloud computing.

1.1 Cloud Computing

Earlier, the mainframe based information infrastructures were developed by the companies which basically have a large, powerful and robust computer to store the data and ran all the required applications. Running the application is easy in mainframe system but their maintenance proved to be expensive and tedious [37].

As the demand of accessing more powerful application prevails, the mainframe computing proved very less effective. The gap was filled with the Distributed computing which involves multiple cheaper systems, each have enough computing resource to run the applications and store the data effectively [2]. But the problem of coordination between the system persist which makes the data sharing difficult.

Cloud computing proved itself as the solution which comes with the advantages of both the above approaches as it offers the central management and coordination as in mainframe computing along with the affordable and scalable feature of distributed computing.

1.1.1 Definition

National Institute of Standards and Technology (NIST) defines cloud computing as a model for enabling universal, convenient, on-demand network access to a shared pool of configurable computing resources like networks, storage, applications, servers and services which can be quickly provided and released with supervision effort and service provider intervention [3].

1.1.2 Characteristics of cloud computing

National Institute of Standards and Technology (NIST) put the following set of characteristics which should hold by any cloud definition. These characteristics help in shifting the business to the cloud [47, 50].

A. **On-demand self service:** client can calculate the computation efforts needed from the cloud.
B. **Broad network access:** services should be available over the internet to access by the client by using smart phones, laptops etc.
C. **Resource pooling:** All the required computing resources should be pooled or managed in such a way that at a time multiple client can access them.
D. **Rapid elasticity:** Computation resources can be increased or decreased by the client on the basis of demand.
E. **Measured services:** The computational services going to be used by the client should be measurable to make them charge pay-per-use basis.

1.1.3 Cloud computing based service models

A. **Infrastructure as a service (IaaS):** Infrastructure as a service proved itself as a most basic level of service provided by cloud computing. In such service, the client do have not bother about the management of the underlying computing resources but have the control

over operating systems, storage, deployed applications and some selected networking components [4].

Amazon web services are an example, which provide infrastructure as a services on a pay per use basis [5].

B. **Platform as a service (PaaS):** In such service of cloud, the client is supposed to deploy their applications using the programming languages and the variety of tools provided by the service provider. The cloud service provider is basically provides a platform where the client do not have the control or can't manage the underlying computing infrastructure such as network, storage and processing etc. but have the control over the deployed application only.

Force.com offered by Sales force is an example of platform as a service [6].

C. **Software as a service (SaaS):** In such kind of service the client is merely the "user" of the application's provided by the service provider. The client needs the browsing application to access such applications which are stored in the cloud.

Google Apps offer such software for public and private entities [7].

1.1.4 Deployment Models

Cloud can be deployed in four ways:

A. **Private Cloud:** This category of deployment model is beneficial for single organization and is specifically used by their employees and solely managed and controlled by an organization or third party. Such models have easier maintenance; high security and organization have much better control [3, 9].

B. **Public Cloud:** As the name suggest, this category is implemented for the general users which is maintained and controlled by cloud services provider organization rather than the user [10]. The users are supposed to pay as per usage. But such systems are more prone to security threats in comparison to the private system as whole data is publicly accessible [11].

C. **Community Cloud:** Deployment of such cloud model is beneficial in the situation where many organizations working with common concern viz. mission, security requirements and policy considerations. In this model, cloud management can be done by one or more organization or also by third party [12].

D. **Hybrid Cloud:** This model is the result of amalgamation of more than one cloud (Private, Public or Community) models. It is bound together by some standard protocols. This helps the organization in resource optimization to increase their core competencies and involved organization can use their private cloud to serve their needs and in case of crucial need they may access the public cloud [3, 13].

2. Underlying Technology

2.1 Virtualization: Virtualization is considered as a very important enabling technology in context of Cloud Computing. It separates resources and services from the physical delivery environment. It helps in hiding the computing resources characteristics from the users. Virtualization actually means "Something which isn't real" but provides the facilities of a real [16].

Virtualization has three main characteristics which make it more useful for cloud computing:

A. **Isolation:** It may be defined as the property, that working of one virtual machine in cloud environment is not going to affect or getting affected by other virtual machine.

B. **Partitioning:** This characteristic is helpful in running many applications and operating system on a single machine.

C. **Encapsulation:** This property helps in storing a virtual machine as a single file.

2.2 Load Balancing: Load Balancing has been emerges out as core technological aspects in cloud computing environment. It is useful in evenly distribution of the processing tasks among the various available processors, for the efficient utilization of the

computing resources which in turn increases the performance and availability of the system [16].

3. Cloud Computing Issues and Challenges

Researchers identify many issues and challenges which are required to be address properly as follows:

A. Security: According to the International Data Corporation (IDC) survey, Security, Availability and Performance are three major issues in adoption of cloud technology. System is more prone to security threats like data loss, phishing etc. because of movement of data, data storage and processing of data outside the control of organization. Security threats can be categorized into two categories – internal and external. Cloud computing also poses privacy concern due to deliberate or accidental access of data which may lead to breach of trust [3, 8- 9]. So novel techniques must be used to reduce the impact of such threats [14].

B. Performance: According to IDC's Survey, Performance is the second major issue in adoption of Cloud Computing Technology. Performance is generally measured in terms of capabilities of application running on the cloud. Poor performance may be caused due to lack of resources like limited bandwidth, less disk space, lower CPU speed and network connections etc. . Low performance may result in low revenue, termination of deal or end of service etc. [15].

C. Reliability and Availability: Reliability and Availability are two major factors for the success of any technology. Reliability gave us idea of availability of the required resources without any failure. Availability can be defined as the time needed by the system to provide the required resources on their demands.

D. Energy Consumption: As per Amazon survey, 53% of total cost is consumed by servers for a 3 years of amortization period while 42% of total budget spent on energy and cooling requirement [15] [27] [28].

4. Literature Review

Many researcher and scientist have proposed the different techniques in area of load balancing in cloud computing environment. Various load balancing technique are discussed below:

In [18], Zhenzhong Zhang et al. proposed an algorithm which used to first forecasts load and then estimates resource requirements of the virtual machines. Experiments put the case that this method can efficiently predict the resource requirements of virtual machines, and also work well with load-balancing framework. Two parameters CPU utilization and memory usage are only taken into consideration.

In [19], Yunhua Deng et al. proposed a Fast Adaptive Load balancing method which uses a binary tree structure for partitioning simulation region into subdomains. The main advantage of this method is faster balancing speed, reduced elapsed time and lesser communication time cost. But Limitations of this method is that it is unable to maintain the underlying topology.

In [20], Kousik Dasgupta et al. proposed the Genetic Algorithm based load balancing technique to minimize the time required by a task sets. As GA works on the concept of natural selection so it provides the reduced response time in comparison other load balancing techniques like First Come First Serve (FCFS), Round Robin (RR), etc. This is very simple approach and same value of mutation and crossover is considered. In this approach, the solutions are represented by Fixed length binary string.

In [21], Ratan Mishra et al. proposed the Ant Colony Optimization (ACO) based load balancing technique. They describe the ant as an unsophisticated insect which have the limited memory and their individual behavior is the main random characteristic of them. Their approach is based on the finding the shortest route which depends on the pheromone traces lay down by ant while travelling from nest to food or vice versa. But this technique has drawbacks which can be rectify by cluster making in clouds.

In [22] Dhinesh Babu L.D, P. Venkata Krishna et al. proposed the Honey bee behavior inspired load balancing technique. In bee hives two types of bees are found i.e. scout bees and forager bees. The scout bees are supposed to find the food sources and pass this news by dancing to propagate an idea about the quantity and quality of food available [23] and then the forager bees visit the location of forged source and to pick the food. Same concept is applied for load balancing of high priority task when it is submitted to virtual machine. The achievement of this method is maximum throughput, minimum waiting time of task and minimum overhead. Drawback is if more high priority task in the queue is available then low priority task have to wait for long time in waiting queue.

In [24], Zhao et al. proposed the link load based load balancing approach. They also discussed the traditional GA is inefficient to meet the objective of load balancing. To balance the traffic load over the network link load problem they used the improved GA and were able to achieve the optimization.

In [25] Yunhua Deng et al. proposed heat diffusion based load balancing which is based on an efficient cell selection scheme. The underlying concept is as follows: when two objects is placed adjacently then heat will transfer from hotter object towards colder object. Similarly, in distributed virtual environment where numbers of users are available and concurrent accessing of processors by number of users can create problem. This situation can be avoided by using this algorithm. According to this algorithm the virtual environment can be divided into large no of square cells and every cell has some objects associated with it. The working is as follow: that every node in cell transfer load to neighboring node in iterative manner, and transfer load is the difference between current node load and load on neighboring node. The advantages of this method are it has less communication overhead, high speed and requires a little amount of calculation. The disadvantages are network delay and wastage of time because of several iterations.

In [31], Wang et al. proposed a task scheduling algorithm which is based on genetic algorithm with the objective of minimizing makespan and evenly distribution of arriving load between

virtual machines. They used greedy algorithm to initialize the population and their selection strategy is based on fitness ratio. Two types of fitness functions are defined, out of which one is selected randomly in each iteration. They have taken adaptive probabilities of crossover and mutation rather than using fixed values.

In [33], Wu G et al. proposed a genetic algorithm based approach for virtual machine consolidation to make data centers energy-efficient. On comparing with existing approaches in which only energy consumed by physical machines is considered, the approach considers energy consumed by both the physical hosts and communication network in the data centers. The solutions produced by the proposed algorithm are 3.5–23.5% better than those produced by FFD [34].

In [36], B. Mondal et al. proposed the novel approach of using the stochastic hill climbing algorithm for load balancing. A Stochastic Hill Climbing is local optimization approach which is used for allocation of incoming jobs to servers or to the virtual machines to get maximum utilization of available resources. A stochastic Hill Climbing and Load optimization algorithm is a loop that continuously moves in one particular direction of rising value which is uphill and stops when it reaches a peak where no neighbor has a higher value in the direction.

PSO has been used diversely in cloud environments. The authors in [48] presented a strategy based on PSO for executing scientific workflows on IaaS clouds.

In [44], Babukarthik et. al. proposed the Hybrid of ACO and Cuckoo search algorithm to solve the task scheduling problem and the algorithm can reduce the total executing time.

Chen et al. [43] presented a workflow scheduling algorithm which is based on Ant Colony System (ACS) algorithm with several new features for its improvement. With the aim to minimize the deadline cost. Three types of heuristic information are defined by them to guide the ants. Out of which, each ant uses one heuristics type and one pheromone type in each iteration which is based on the probabilities controlled by the parameters.

In [49], Shojafar et. al. introduced a FUGE i.e. hybridization of GA and fuzzy logic that intends to minimize makespan, degree of imbalance and cost in cloud. In this, technique two different

types of chromosomes are created which are based on diverse QoS parameters. Fuzzy theory is used for the computation of the fitness value and for crossover operation. The proposed approach achieves better performance when compared with ACO.

5. Motivation

Cloud computing evolved from the adoption of existing technologies with the aim of less cost and ease of resource management. Virtualization is the underlying technology in cloud computing which provides the activeness in operations and also provides the maximum utilization of computing resources. The main goal of virtualization is its ability to run various virtual machines on a single machine because of sharing of the available computing resources. The main challenge before them is to maintain the existing or better level of service during the period of outburst in demand. Thus in spite of golden future of cloud computing, still may critical issues need to be tackled for its perfect realization [8]. Load balancing is also one of the major issues. The problem of load sharing occurs when the several jobs from the users trying to access the same server or VM's while the other are idle as of no request from the users. This situation is known as distributed load imbalance system. The problem can be addressed by scheduling the request or tasks before they came to the server.

As load balancing in a distributed environment like cloud, grid etc. is very crucial. Efficient Load balancing techniques allows distribution of workloads effectively across the available datacenters, servers, hard drives and other computing resources. Job scheduling or load balancing is a combinational problem. It cannot be considered as linear programming, and it is well known as NP-complete problem. In order to solve such problem, various approximation and branch and bound methods are proposed, but the result is unsatisfactory and needs a lot of time for computation that is not practical in cloud environment. Moreover, the goal of cloud computing can be too complex and which depends on the orientation of business in cloud environment, which makes it impossible to get solved in linear time by using the existing traditional scheduling

algorithm [42]. As energy consumption by data centers has become a major issue, a lot of research is going on in the direction of energy aware scheduling. A conventional approach to save energy in data centers is to perform virtual machine consolidation that is virtual machines are packed on minimum physical machines.

6. Soft Computing Techniques

In real world, there are many problems which we cannot solve logically, or we can say some problems can be solved theoretically but practically impossible due to huge resource and computational time requirements [39]. For such kind of problems, sometimes the biological natural methods may prove efficient and effective. As a near optimal solution is sometime more beneficial in most practical problem. Such biological inspired methods are called Soft Computing Techniques. Soft computing differs from the conventional hard computing as it is tolerant of uncertainty, imprecision, low solution cost, approximation and robustness.

Basic technical disciplines of soft computing are Fuzzy logic (FL) [26], Probabilistic Reasoning (PR), Neural Computing (NN), and Evolutionary Computing (EC):

6.1 Genetic Algorithm

Genetic Algorithms (GA) was introduced by John Holland in 1975. These are searching algorithms which are basically based on the mechanics of natural genetics, the method which drives biological evolution, such as inheritance, mutation, selection and crossover. They belongs to the higher class of Evolutionary algorithms (EA) They use iterative approach which consists of constant size population of candidate solution called individuals, creatures, phenotype. Every individual is represented by finite string of symbols. The symbols are generally 0's and 1's (binary), although other representations like character based encodings, real-value encodings and tree representation are also permitted. The process of evolution usually begins from the population of randomly generated individuals, follows an iterative process, the population in each

iteration is known as generation. In each generation, a fitness score is assigned to each solution which represents the ability of an individual to compete. The GA maintains the population of n solutions with fitness value. Parents are the most fit individuals are allowed to mate to form a new generation and this new generation is then used in next level of iteration [38]. As this iterative process goes on, the algorithm terminates once the population has converged i.e. not producing the offspring not different noticeably from the previous offspring's [29] [30].

The operating rules of GA are as follows:

A. Selection – as per concept of survival for fittest, parents are selected.

B. Crossover – parents are open to mate, creation of next generation.

C. Mutation – making random changes in the parents to form children.

6.2 Multi-Objective optimization

Many optimization problems naturally have several or multiple objectives which need to be satisfied but these objectives may be conflicting in nature, but in order to find the solution of such problem, they may be treated as they only one such objective while the other may be handled as constraints [35].

Evolutionary multi-Objective optimization can be classified as follows:

A. Non-Pareto techniques – in this approach the multiple objectives can be converted into a single one by using some weight function or using one for the optimization while the other transform into a constraint. But such method may result in huge loss of significant trade-off and unable to find a global optimum. Techniques under such category are : Vector Evaluated Genetic Algorithm (VEGA), Aggregate approaches, Lexicographic ordering etc.

B. Pareto based techniques – in this approach, solution that presents optimal trade-offs is selected. The main goal is to find the non-dominated global solutions. Such set of non-dominated solutions are known as Pareto-Optimal set which are named after Vifredo Pareto [Fonseca and Fleming 1995]. Techniques under such category are: Multi-

Objective Genetic Algorithm (MOGA) proposed by Fonseca and Fleming, Non-Dominated Sorted GA (NSGA) by Srinivas and Dev, and NSGA-II.

6.3 Motivation for Soft Computing Techniques

Task scheduling in cloud computing environment may be defined as the mapping of n jobs or tasks to m processing units or machines. So there are m^n ways for the allocation of jobs to such machines. Task scheduling is an important issue, used to greatly influence the performance of cloud computing environment. In the real time environment, allocation of tasks to the resources is very complex problem and therefore, complexity class of the task scheduling problem is categorized as NP-complete because of extremely large search space with correspondingly large number of potential solutions [54, 56] and it takes huge time to get the optimal solution. Load balancing issue in dynamic task scheduling is an NP-hard problem [55] as task migration may be required for the processing of task due to non-availability of machine. Many researchers treated task scheduling as a single objective optimization problem, such as minimization of make span and solved conventionally. But scheduling decisions may require minimization of two or more conflicting objectives instead of single objective. Solving a multi-objective scheduling problem is more complicated than solving a single objective scheduling problem.

To solve NP-complete problems, evolutionary and heuristic algorithm are used, and for distributed systems scheduling problems algorithms like Particle Swarm Optimization (PSO), Simulated Annealing, Tabu Search and Genetic Algorithm (GA) are used. In heuristic Min-Min and Max-Min algorithms that are begun by a batch of unmapped tasks and a task's execution time on every processor [52] is computed. Ant Colony Optimization algorithm solves optimization problems to find good paths. The aim is to achieve minimum weighted flow time for jobs and minimizing makespan [53].

With the increase in number of user's size for associated computing, sometime the tasks which needed to be scheduled are proportionally increases, and the existing conventional task scheduling strategies cannot fulfill its requirements. . Optimal task scheduling to minimize the

sum of task running and communication costs using the branch-and-bound technique and by using simulation methods, we can calculate computational complexity of this method [43]. Conventional methods can apply in optimization are deterministic, fast, and give exact answers but often tends to get stuck on local optima [44]. For the above reason better algorithms for task scheduling is always needed to reduce computation time and the associated cost [51]. An efficient task scheduling algorithm directly affects the system performance and soft computing based techniques proves to be very effective from the scheduling point of view [49, 44]. In cloud computing environment, it is more than sufficient to find near optimal solution, if possible in a short period of time. In this situation heuristic approach plays significant role. Meta heuristic methods for scheduling like genetic algorithms (GA), simulated annealing (SA), particle swarm optimization (PSO), and ant colony optimization (ACO) are used to solve the well-known NP-Complete problems. Other soft computing techniques like fuzzy logic, Quantum GA, Neural Network with their integration with meta-heuristic may give better results.

7. Proposed work

Efficient resource management can be considered as a prerequisite in cloud computing for the maximum or efficient utilization of computing resources in distributed computing environment which results in economic investment from the client's point of view. Load balancing or task scheduling in cloud computing environment belongs to a category of NP-hard problem because of the availability of large solution space and it takes huge time to reach to an optimal solution. Load balancing in cloud computing consists of n tasks and m Virtual Machines (VM's). Each task should be processed by each of the m VM's in order to minimize makespan. Scheduling algorithms are used mainly to minimize execution time and execution cost [40]. The Task Scheduling handles the problem in which resources need to be assigned to the received job. An efficient scheduling algorithm considers balancing the load of the system considering the total execution time of available resources, leading to minimization of execution cost.

Following objectives are set for this research work to accomplished:

A. Identification and analysis of the existing techniques for efficient load balancing or task scheduling in cloud computing environment.

B. To develop the efficient task scheduling mechanism to achieve the resource optimization in cloud computing environment using the best suitable soft computing techniques like genetic algorithm, particle swarm optimization, ant colony optimization, quantum GA, fuzzy logic etc.

C. To explore the possibilities using the integration of soft computing techniques for the development of efficient task scheduling mechanism.

D. Evaluation and comparison of the developed task scheduling techniques.

References

[1] McCarthy, J. (1961). Speech given at MIT. Time-Sharing Computer Systems.

[2] Foster, I., Zhao, Y., Raicu, I., Lu, S. (2008) “Cloud Computing and Grid Computing 360-Degree Compared”, University of Chicago, 10-56.

[3] Badger, L., Grance, T., Patt-Corner R., Voas, J. (2012) "Cloud Computing Synopsis and Recommendations", National Institute of Standards and Technology, Special Publication 800-146, http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf

[4] Buyya, R., Yeo, C., Venugopal, S., Broberg, J., Brandic, I. (2009) “ Cloud Computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility”. Future Generation Computer Systems, 25, 599- 616.

[5] About Amazon Web Services. Amazon Web Services Homepage. Available at: http://aws.amazon.com/. Accessed on: June 2015.

[6] Salesforce (2010). Force.com Product. Available at: http://www.salesforce.com/platform/. Accessed on: April 2015.

[7] About Google Apps. Available at: http://www.google.com/apps/intl/en/business/index.html. Accessed on: April 2015

[8] Vouk, A. (2008) “Cloud computing-issues, research and implementations”, in Proceedings of the ITI 2008 30th Int. Conf. on Information Technology Interfaces.

[9] Dillon, T., Wu, C. and Chang, E. (2010) “Cloud Computing: Issues and Challenges”, 24th IEEE International Conference on Advanced Information Networking and Applications, IEEE Computer Society.

[10] Pedram, M. (2012) “Energy-Efficient Datacenters”, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, Vol. 31, NO. 10, pp. 1465-1484

[11] IBM Global Services. (2009) “Cloud computing: defined and demystified explore public, private and hybrid cloud approaches to help accelerate innovative business solutions”.

[12] Sakr, S., Liu, A., Batista, D.M., Alomari, M. (2011) “Survey of Large Scale Data Management Approaches in Cloud Environments”, IEEE Communications Surveys & Tutorials, Vol. 13, No. 3, Third Quarter, pp. 311-336.

[13] Dikaiakos, M.D., Pallis, G., Katsaros D., Mehta, P., Vakali, A. (2010) “Cloud Computing: Distributed Internet Computing for IT and Scientific”, 1089-7801/09 IEEE Computer Society.

[14] Kuyoro, S. O., Ibikunle,F., Awodele, O. (2011) “Cloud Computing Security Issues and Challenges”, International Journal of Computer Networks (IJCN), 3(5): pp. 247-255.

[15] Sajid, M. and Raza, Z. (2013) “Cloud Computing: Issues & Challenges”, in International Conference on Cloud. pp. 35-41.

[16] MacVittie, L. (2009). “Load balancing is key to successful cloud-based (dynamic) architectures”, DevCentral Home. Available at: http://devcentral.f5.com/weblogs/macvittie/archive/2009/01/23/loadbalancing-is-key-to-successful-cloud-based-dynamic-architectures.aspx. Accessed on: July 2015.

[17] Anthony, T., Toby J.V., Elsenpeter, R. (2010) “Cloud Computing a Practical Approach”, TATA McGraw-Hill Edition.

[18] Zhenzhong, Z., Limin X., Yuan T., Tian J., Shouxin W., Hua L. (2013) “A Model Based Load Balancing Method In IaaS Cloud”,42nd international conference on parallel processing.

[19] Yunhua D., Rynson W.H. L., (2010) “Heat diffusion based dynamic load balancing for distributed virtual environments”, in: Proceedings of the17th ACM Symposium on Virtual Reality Software and Technology, ACM.

[20] Dasgupta, K., Mandal, B., Dutta, P., Mondal, J.K., Dam, S. (2013) “A Genetic Algorithm (GA) based Load Balancing Strategy for Cloud Computing”, in Proc. of Elsevier, Procedia Technology.

[21] Mishra, R., Jaiswal, A. (2012) “Ant colony Optimization: A Solution of Load balancing in Cloud”, in International Journal of Web & Semantic Technology, Vol.3, No.2, pp. 33-500.

[22] Babu, D., VenkataKrishna, P. (2013) “Honey bee behavior inspired load balancing of tasks in cloud computing environments”, Applied Soft Computing, ASOC 1894 1–12, Elsevier B.V.

[23] Zhao, L., Dong, Y.Min., Haung, C.Y. (2013) “Study of link load balancing based on improved Genetic Algorithm”, IEEE COMPUTER SOCIETY.

[24] Chaczko, Z., Mahadevan, V., Aslanzadeh, S., Mcdermid, C. (2011) “Availability and Load Balancing in Cloud Computing”, International Conference on Computer and Software Modeling.

[25] Deng, Y., WH Lau, R. (2010) “Heat diffusion based dynamic load balancing for distributed virtual environments”, in: Proceedings of the17th ACM Symposium on Virtual Reality Software and Technology, ACM.

[26] Zadeh, A. (1994), “Fuzzy Logic, Neural Networks and Soft Computing”, Communication of the ACM, 37(3), pp77-84.

[27] Energy Star, Report to congress on server and data centre energy efficiency, Public Law 109-431, U.S. Environmental Protection Agency, August 2007.

[28] Hamilton, J. (2009) “Cooperative expendable micro-slice servers (CEMS): Low cost, low power servers for internet-scale services”, Proc. 4th Biennial Conf. Innovative Data Systems Research (CIDR).

[29] Moraga, R.J., DePuy, G.W., Whitehouse, G.E. (2006) “Metaheuristics: a solution methodology for optimization problems”, Handb Ind Optim Probl Handb Ind Syst Eng AB Badiru. http://dx.doi.org/10.1201/9781420038347.

[30] Pop, F., Dobre, C., Cristea, V. (2009) "Genetic algorithm for DAG scheduling in grid environments", In: 5th IEEE Int Conf intell comput commun process. p. 299–305.

[31] Wang, T., Liu, Z., Chen, Y., Xu, Y., Dai, X. (2014) "Load balancing task scheduling based on genetic algorithm in cloud computing", In: IEEE 12th International Conference on dependable auton secur comput, p. 146–52.

[32] Joseph, C.T., Chandrasekaran, K., Cyriac, R. (2015) "A novel family genetic approach for virtual machine allocation" Proc Computer Science ; 46:558–65.

[33] Wu, G., Maolin, T., Tian, Y.C., Li, W. (2012) "Energy-efficient virtual machine placement in data centers by genetic algorithm", In: Vmslv A, editor. Neural inf process. Springer, p. 315–23.

[34] Kalra, M., Singh, S. (2015) "A review of metaheuristic scheduling techniques in cloud computing", Egyptian Informatics J, http://dx.doi. org/10.1016/j.eij.2015.07.001

[35] Deb, K., (2001) "Multi-Objective Optimization Using Evolutionary Algorithms", Chichester, U.K.: Wiley.

[36] Mondal, B., Dasgupta, K., Dutta, P. (2012) "Load Balancing In Cloud Computing using Stochastic Hill Climbing-A Soft Computing Approach" Science Direct C3IT

[37] Mezmaz, M., Melab, N., Kessaci, Y., Lee, Y.C., Talbi, E.G., Zomaya, A.Y. and Tuyttens, D. (2011) "A Parallel Bi-Objective Hybrid Metaheuristic for Energy-Aware Scheduling for Cloud Computing Systems", Journal of Parallel and Distributed Computing, 71, 1497-1508.

[38] Moraga, R.J., DePuy, G.W., Whitehouse, G.E. (2006) "Metaheuristics: a solution methodology for optimization problems", Handb Ind Optim Probl Handb Ind Syst Eng AB Badiru. http://dx.doi. org/10.1201/9781420038347.

[39] Soft Computing. https://en.wikipedia.org/wiki/Soft_computing: Accessed on June, 2015

[40] Tsai, J.T., Fang, J.C. and Chou, J.H. (2013) "Optimized Task Scheduling and Resource Allocation on Cloud Computing Environment Using Improved Differential Evolution Algorithm. Computers & Operations Research", 40, 3045-3055. http://dx.doi.org/10.1016/j.cor.2013.06.012

[41] Dutta, D. and Joshi, R.C. (2011) “A Genetic-Algorithm Approach to Cost-Based Multi-QoS Job Scheduling in Cloud Computing Environment”, International Conference and Workshop on Emerging Trends in Technology (ICWET 2011) TCET, Mumbai.

[42] Rimal, B. P., Eunmi, C., Lumb, I. (2009) “A.: Taxonomy and Survey of Cloud Computing Systems”, In: Fifth International Joint Conference on INC, IMS and IDC, Seoul, pp. 44–51.

[43] Chen, W.N., Zhang, J.Z.J. (2009) “An ant colony optimization approach to a grid workflow scheduling problem with various QoS requirements”, IEEE Trans Syst Man Cybern Part C 29–43.

[44] Babukarthik, R.G., Raju, R., and Dhavachelvan, P. (2013) “Hybrid Algorithm for Job Scheduling: Combining the Benefits of ACO and Cuckoo Search”, in Advances in Computing and Information Technology. Springer Berlin Heidelberg, pp. 479-490.

[45] Dorigo, M., Stutzle, T. (2006) “Ant colony optimization”, in IEEE Computational Intelligence Magazine, pp. 28-39.

[46] Sadiku, M., Musa, S., Momoh, O. (2014) “Cloud computing: opportunities and challenges”, IEEE Potentials **33 (1)** 34–36.

[47] Duan, Q., Yan, Y., Vasilakos, A.V. (2012) “A survey on service-oriented network virtualization toward convergence of networking and cloud computing”, IEEE Trans. Netw. Service Manage**. 9 (4)** 373–392.

[48] Rodriguez, S. M., Buyya, R. (2012) “Deadline based resource provisioning and scheduling algorithm for scientific workflows on clouds”, IEEE Trans Cloud Computing ; 2:222–35.

[49] Shojafar, M., Javanmardi, S., Abolfazli, S., Cordeschi, N. (2015) “FUGE: A joint meta-heuristic approach to cloud job scheduling algorithm using fuzzy theory and a genetic method”, Cluster Comput 2015:1–16. http://dx.doi.org/10.1007/s10586-014-0420-x.

[50] Abbas, A., Bilal, K., Zhang, L., Khan, S.U. (2014) “A cloud based health insurance plan recommendation system: a user centered approach”, Future Gener. Comput. Syst., http://dx.doi.org/10.1016/j.future.2014.08.010.

[51] Guo, L., Zhao, S., Shen, S. and Jiang, C. (2012) "Task Scheduling Optimization in Cloud Computing Based on Heuristic Algorithm", Journal of Networks, Vol. 7, No. 3, 547-553.

[52] Preethima, R.A., Margret, J. (2013) “Survey on optimization techniques for task scheduling in cloud environment”, Interna-tional Journal of Advanced Research in Computer Science and Software Engineering. 3(12):413–5.

[53] Umarani, G., Uma, V., Shanthi, A.P., Arul, S. (2015) “Task scheduling model”, Indian Journal of Science and Technology. 8(S7):33–42.

[54] Garey, M.R. & Johnson, D.S. (1979) “Computer and Intractability, a guide to the theory of NPCompleteness”, W.H. Freeman and Company, ISBN:0-7167-1044-7, New York.

[55] Brucker, p. (2001) “Scheduling Algorithms”, Springer, Berlin 3rd edition.

[56] Ullman, J.D. (1975) “Np-complete scheduling problems”, J.Comput. Syst. Sci, 10(3).